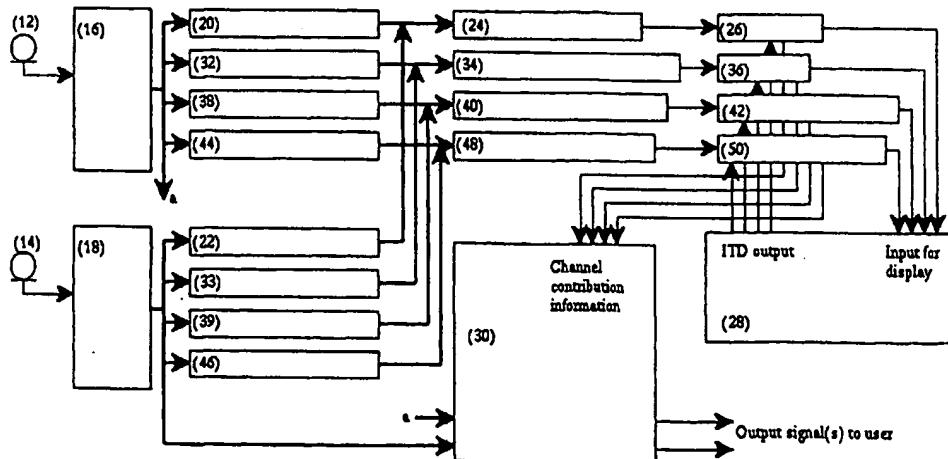




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> :  H04R 25 /00		A1	(11) International Publication Number: <b>WO 00/01200</b>  (43) International Publication Date: 6 January 2000 (06.01.00)
<p>(21) International Application Number: <b>PCT/GB99/02063</b></p> <p>(22) International Filing Date: 30 June 1999 (30.06.99)</p> <p>(30) Priority Data: 9813973.6 30 June 1998 (30.06.98) GB</p> <p>(71) Applicant (for all designated States except US): UNIVERSITY OF STIRLING [GB/GB]; Stirling FK9 4LA (GB).</p> <p>(72) Inventor; and</p> <p>(73) Inventor/Applicant (for US only): SMITH, Leslie, Samuel [GB/GB]; 1 Beech Court, Dunblane FK15 0LA (GB).</p> <p>(74) Agents: McCALLUM, William, Potter et al.; Cruikshank &amp; Fairweather, 19 Royal Exchange Square, Glasgow G1 3AE (GB).</p>		<p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>	

(54) Title: METHOD AND APPARATUS FOR PROCESSING SOUND



(57) Abstract

A method of processing sound comprises the steps of: detecting sounds at at least two spaced detecting locations; analysing the detected sounds to identify the angular relation between respective sound sources and the detecting locations; permitting selection of an angular relation associated with a particular sound source; and processing the detected sounds in response to the selection to highlight a stream of sound associated with the particular sound source. The method may be utilised in a hearing aid, to allow a user to stream sounds by interactively selecting a particular source, and thus minimise background "noise".

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD AND APPARATUS FOR PROCESSING SOUNDFIELD OF THE INVENTION

The present invention relates to method and apparatus for processing sound, and in particular but not exclusively to a hearing aid, and in particular to an interactive directional hearing aid.

5      BACKGROUND

Most current hearing aids tackle the problem of hearing loss by (i) detecting the sound using a single microphone, (ii) selectively transforming the incoming sound, possibly initially converting the sound to a digital form so that more sophisticated digital signal processing techniques can be used, and (iii) re-transmitting the sound in the ear canal (or, in the case of cochlear implants, directly stimulating the nerves of the spiral ganglion in the organ of Corti). The use of a single microphone means that selectively amplifying sounds coming from a particular direction can only be achieved by taking advantage of the shape of the directional response of the microphone. However, using a highly directional microphone leads to a different problem: the inability to detect sounds from certain other directions. In embodiments of the present invention, we use two microphones and incorporate methods apparently used by animal auditory systems to separate out

the different sources (or streams) of sound present in incident sound.

Even where two or more microphones are used, few existing systems permit the user to interact with the 5 system to select the most appropriate sound processing. This is a disadvantage because the nature of the problem the user faces in sound interpretation depends strongly on the user's environment; this may vary from a quiet environment, with only one source of sound, to a noisy 10 room, with many different sources of sound. The primary information used by the auditory system for determining the direction of a sound source is interaural intensity differences (IIDs), and interaural time differences (ITDs). To be able to estimate IID or ITD, a hearing aid must have 15 more than one microphone.

US patent 3946168, and US patent 3975599 disclose use of two microphones in a single housing, pointing in different directions, and switched between the directional input signals, essentially taking advantage of the 20 different directional characteristics of the two microphones. A similar approach, using both omni- and unidirectional microphones and including some adaptive equalisation is taken in US patent 5524056. A more sophisticated approach (US patent 4751738) uses a number of 25 microphones in pairs, spaced one half-wavelength (of the frequencies of interest) apart across the user's body. The signals from these microphones are summed, bandpassed, and amplified. This provides directionality in the region of

the chosen frequencies in the direction that the user is facing. This is extended in US patent 5737430 to include wireless connection to an ear-placed hearing aid.

The advent of portable digital signal processing (DSP) has meant that more sophisticated signal processing strategies can be adopted. DSP techniques have been applied to binaural systems (which have two microphones and two output transducers, one per ear) (US patent 5479522, US patent 5651071), resulting in a system which selectively amplifies signals characteristic of speech, while maintaining the precise timing of the signals so as to permit the user to detect sound source direction. In this way the systems also perform noise reduction. Directionality has been added using DSP techniques to perform beamforming (US patent 5511128). Implementation of these entities using wireless communication is described in US patent 5757932. Techniques which attempt to compromise between the conflicting goals of maximally directional response, and preservation of IID and ITD binaural cues for source direction finding are compared by Desloge et al (J.G. Desloge, W.M Rabinowitz, and P.M. Zurek, Microphone-array hearing aids with binaural output - part 1:fixed-processing systems. IEEE Transactions on Speech and Audio Processing, 5(6):529--542, 1997). In Kollmeier et al (B. Kollmeier, J. Peissig, and V. Hohmann. Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain. Scandinavian Audiology: Supplement, 38:28--28, 1993), an algorithm which attempts to amplify

only those sounds with the appropriate IID and ITD for sources straight ahead is described.

Another multiple microphone technique pioneered at the University of Paisley, Scotland uses sub-band adaptive processing, and this allows statistically different signals (such as speech and car noise) to be separated, improving the signal-to-noise ratio (SNR) (P.W. Shields and D. Campbell, Multi-microphone sub-band adaptive signal processing for improvement of hearing aid performance: preliminary results using normal hearing volunteers; Proc ICASSP97, pages I, 415--418, 1997; P. Shields, M. Girolami, D. Campbell, and C. Fyfe, Adaptive processing schemes inspired by binaural unmasking for enhancement of speech corrupted with noise and reverberation; in L.S. Smith and A. Hamilton, editors, *Neuromorphic Systems: engineering silicon from neurobiology*, pages 61--74. World Scientific, 1998; A Hussain and D. Campbell, Binaural sub-band adaptive speech enhancement using a human cochlear model and artificial neural networks, in L.S. Smith and A. Hamilton, editors, *Neuromorphic Systems: engineering silicon from neurobiology*, pages 75--86. World Scientific, 1998). Additionally, anti-Hebbian learning techniques from the blind signal deconvolution (independent components analysis) school can be used, allowing different sound streams to be recovered.

Both normal and hearing-impaired listeners can move their heads to assist in picking out the source in which they are interested. The earliest hearing aids were

mechanical, and the user could move them to alter the direction in which they pointed. Most early electronic hearing aids had little additional user interfacing: they could be switched on or off, and had, perhaps, a number of settings, and/or a volume control. Most modern hearing aids are configurable, and are set at dispensing to compensate for the particular hearing loss of the user: however, they are not usually user-reconfigurable thereafter.

10 Some recent hearing aids have additional alterable settings. In US patent 5524056, the particular microphones to be used may be altered. Similarly in US patent 5636285 a voice-controlled re-settable technique is described.

15 Most hearing aids retransmit sound in the auditory canal. The provision of appropriate selective transformations between the incoming sound and the transmitted sound is normally based on making up the hearing that the user appears to have lost. However, partial deafness is not simply a decrease in sensitivity in parts of the spectrum: if it were, then this approach would be entirely successful. The problem is that much of the loss of sensitivity is due to failure of the hair cell transduction system, particularly at the basal (high frequency) end of the cochlea. Simply amplifying high frequency signals will not result in stimulation of the auditory nerve cells that these inner hair cells innervate. Instead, other (undamaged) hair cells from elsewhere in the cochlea will respond, mixing their response to the

amplified high frequency sounds with their original response. Selective amplification thus results in increases in the response on the auditory nerve to a wider range of the spectrum, but at the expense of place-based frequency resolution. Further, for many hearing impaired subjects, the distance between audible sounds and painful sounds is small, making the procedure of adjusting a frequency-sensitive amplifier difficult.

For subjects with little or no residual hearing, cochlear implant techniques are often used in place of auditory retransmission. These excite the neurons of the spiral ganglion directly.. Unfortunately, it is not possible to stimulate all of the auditory nerve in this way, as the shape of the cochlea precludes this. Thus, only the high frequency (basal) end of the cochlea can be stimulated so that the user is presented with a much impoverished signal.

Another possibility when there is little or no residual hearing, (and in particular where there is damage to the auditory nerve or brainstem) is to use a different modality, such as the visual modality. This is suggested in US patent 5029216, where a spectacle-mounted system which can give warning to a hearing-impaired driver that there is an emergency vehicle approaching is described.

Whether the sound is retransmitted, whether the auditory nerves are directly stimulated, or whether the visual domain is used, both the sounds of interest and noise are likely to be presented. One can concentrate on

areas of the spectrum in which speech is most likely to occur, but then, one will be amplifying all the speakers talking at once, leading to the commonly found problem that users of hearing aids can pick out speech when one speaker talks, but not when there are a number of speakers. Separating out what is of interest and what is noise is difficult because it is likely to vary in time and as the listener moves around. The result is that the signal in which the user is interested and noise tend to both be amplified. It was these problems that the multi-microphone techniques discussed above aimed to solve: however, their directionality is often restricted to the direction in which the user is facing.

The physiology of the early auditory system is well known, and well described in, for example, J.O. Pickles, An Introduction to the Physiology of Hearing; Academic Press, 2nd edition, 1988. This physiology is very similar across a wide range of mammals and this suggests that whatever is happening at this stage is (i) effective, and (ii) not predicated on specifically human aspects of auditory processing. We suggest that what is going on is that the sound is being streamed (A.S. Bregman. Auditory scene analysis MIT Press, 1990) both monaurally and binaurally. This seems likely (i) because the same problems of streaming are found across the animal kingdom and (ii) because logically, streaming of sounds should precede interpretation.

The auditory system appears to use a number of

different cues in performing streaming. For binaural streaming, these include relative intensity, relative timings of sudden increases in intensity in parts of the spectrum (onsets), relative timing of features of the envelope of bandpassed sound (most notably amplitude modulation peaks), and relative timings of the peaks and troughs of the bandpassed signal (waveform-synchronous features). Relative intensity is most used at low frequencies where the head shadow results in sounds from the sides being much stronger in one ear than the other: this is less pronounced at higher frequencies due to the sound waves diffracting round the head. For monaural streaming, the co-occurrence and relative timing across the spectrum of onsets, and the co-occurrence of same-frequency amplitude modulation in medium and high frequency areas of the spectrum appear to be used. This list is not intended to be exhaustive, but to give examples of the range of techniques in simultaneous use by the early auditory system.

Apart from relative intensity, all the features above have their roots in the fine time-structure of the sound. These features may be grouped into three classes (S. Rosen. Temporal information in speech: acoustic, auditory and linguistic aspects. Phil. Trans. R. Soc. London B, 336:367-373, 1992; L.S. Smith. Extracting features from the short-term structure of cochlear filtered sound. In J.A. Bullinaria, D.W. Glasspool, and H. Houghton, editors, 4th Neural Computation and Psychology Workshop, London, 9-

11 April 1997, pages 113--125. Springer Verlag, 1998) and features from all three classes contain information which may be used in monaural grouping and sound direction finding.

5       The primary source of the differences in these features between the two ear systems are the inter-aural time difference (ITD) and inter-aural intensity difference (IID). The exact forms that these take are described in for example J. Blauert, Spatial Hearing, MIT Press, revised  
10      edition, 1996. From this discussion, it is clear that IID is more effective at low frequencies (due to the shadow effect of the head), and ITD at medium and high frequencies (because the signal period is large compared with the difference in signal path times to each of the two ears)  
15      (see figure 1).

Because of the way in which the inner hair cells of the organ of Corti transduce the pressure wave on the basilar membrane of the cochlea, they are more likely to cause a spike on the auditory nerve at one part of the  
20      phase of the incoming signal, at least for frequencies between about 20Hz and 4KHz. This phase locking is believed to be important in the detection of the inter-aural time difference later in the processing of auditory signals. So long as the period of the signal is long compared to the  
25      ITD, this provides an important and unambiguous cue. However, at, for example, 2.5 KHz, the period is 400 microseconds. If the ITD is 150 microseconds (corresponding to  $\theta=25$  degrees), then we would expect the

signals reaching the two ears to be  $3\pi/4$  out of phase. But this is not distinguishable from the signals being  $2\pi - 3\pi/4$  out of phase, corresponding to an ITD of 250 microseconds, which is  $\theta = 37$  degrees. So, medium to high frequencies result in ambiguous directions.

At high frequencies (above about 4KHz), the phase-locking breaks down, so that waveform synchronous ITD cannot be used to locate constant high frequency sounds.

Sudden increases in intensity of sound result in onset cells in the cochlear nucleus firing (J. O. Pickles. An Introduction to the Physiology of Hearing. Academic Press, 2nd edition, 1988.). These onset cells have a short latency, and are relatively insensitive to intensity (J.S. Rothman, E.D. Young, and P.D. Manis. Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: Implications of a computational model. *Journal of Neurophysiology*, 70(6):2562--2583, 1993; J.S. Rothman and E.D. Young. Enhancement of neural synchronization in computational models of ventral cochlear nucleus bushy cells. *Auditory Neuroscience*, 2:47--62, 1996). Since the intensity is likely to be different at the two detectors this is important. The latency is very short, and the population coding of a number of cells is believed to permit the timing of the onset to be precisely measured (D.C. Fitzpatrick, R. Batra, T.R. Stanford, and S. Kuwada. A neuronal population code for sound localization, *Nature*, 388:871-874, 1997; B.C. Skottun. Sound localization and neurons. *Nature*, 393:531, 1998). As a

result, the ITD of the onset envelope can provide useful directional information, even for signals at high frequencies.

In a normal environment, most sounds reach the ears by many different paths, due to the presence of reflecting surfaces. One result of this is that the ITD is the result of the combination of these multiple paths, and this may cause incorrect estimates of the direction of the sound source. However, the direct path is always the fastest, and generally the least attenuated. Thus, the sound direction computed from initial onsets is not affected by the existence of multiple paths. Additionally, onsets generated by the arrival of signals from reflected paths are attempting to generate responses from the same onset cells that just fired because of the signal from the direct path. These reflected onsets will not be as strong as the original onset, and will be attempting to stimulate cells which are likely to be in their refractory period.

Many pitched sounds consist of multiple harmonics of a low-frequency fundamental. This is true of many animal noises, including voiced sounds in speech. As a result, being able to find the direction of these is particularly important. The nature of the bandpass filtering which occurs on the cochlea is such that for these sounds, a number of adjacent harmonics are present at many higher frequency locations of the cochlea. This results in the energy of the movement of the basilar membrane of the cochlea being modulated in amplitude at the frequency of

the fundamental. The result is that the auditory nerve output is similarly modulated. Stellate (chopper) cells in the cochlear nucleus appear to be particularly sensitive to amplitude modulated signals, and can amplify this modulation (A.R. Palmer and I.M. Winter. Cochlear nerve and cochlear nucleus responses to the fundamental frequency of voiced speech sounds and harmonic complex tones. *Advances in the Biosciences*, 83:231--239, 1992). Using the difference in the timings of the peaks and troughs of this amplitude modulation, the auditory system can find the direction of certain high frequency sounds, even although waveform-synchronous phase locking is absent.

Recent work in auditory psychophysics suggests that the ITDs detected from onsets, amplitude modulation, and waveform synchronous processing are not used directly, but are grouped monaurally first (S. Carlile, *The Physical and Psychophysical Basis of Sound Localization*, in *Virtual Auditory Space: Generation and Applications*, edited by S. Carlile, R. G. Landes Company, 1996; J. F. Culling and Q. Summerfield, *Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay*, *J. Acoustical Soc. of America*, 98, Vol. 2, 785-796, 1995; C. J. Darwin and V. Ciocca, *Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component*, *J. Acoustical Soc. of America*, 91, 6, 3381-3390, 1992; C. J. Darwin and R. W. Hukin, *Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity*, *J.*

Acoustical Soc. of America, 102 (4), 2316-2324, 1997).

This appears to be most effective because the ITDs are computed from a number of channels simultaneously, reducing the likelihood of errors.

5        Most real environmental sounds are complex, containing energy at many different frequencies. Some are unpitched, and some are pitched. But we do not normally notice any particular difficulty in determining the direction of different types of sound. We suggest this is because the 10      auditory system uses all of the techniques above, plus IID (and perhaps some other techniques of which we are not aware). Certainly, (i) IID is useful particularly with low frequency sounds (ii) onsets are useful with sounds which start suddenly, whether pitched or unpitched (such as a handclap) (iii) waveform-synchronous techniques are useful with medium frequency sounds and (iv) amplitude modulation based techniques are useful with high frequency sounds which display amplitude modulation when bandpass-filtered. (Note that people find it difficult to find the precise 15      direction of pure constant high frequency tones: the IID is small, and the waveform synchronous processing breaks down).

20     

25      Hearing loss causes loss of information on all aspects of the fine time structure. Clearly for those bands of the signal not detected at all at the inner hair cells there will be no fine timing information available at all. Additionally, for those bands of the signal for which an area of the organ of Corti is non-functional, detection

will occur only on neighbouring areas of the organ of Corti. There may be a loss of synchronisation of the auditory nerve signal to both the features of the envelope modulation, and to the peaks and troughs of the bandpassed signal itself. We suggest that this loss of fine timing information is one of the primary reasons for hearing impaired people finding sound streaming difficult.

An object of the embodiment of the present invention is to provide an interactive system whereby the classes of features of sound may be detected synthetically, and used to find the direction of incoming sounds. This directional information can then be used to stream sounds.

#### SUMMARY OF THE INVENTION

Accordingly, the preferred embodiment of the present invention provides an interactive system in which the precise timing of the signals produced by bandpassing incoming sounds at two microphones is detected using techniques based on what appears to happen in the early auditory system. This, along with the IID in each channel, is used to determine the direction of the sound source, and this directional information is displayed to the user. The user selects which elements of the sound should be presented by interactively selecting some of the elements displayed. User interaction may consist of the user pointing their head in a particular direction, or may take place using some form of display and graphics tablet. The final presentation of the selected auditory information may

use the auditory modality (using selective amplification, attenuation and possibly resynthesis), or the visual modality.

5 The system described here in may also be used as a part of an auditory system for a robot. Sound sources from particular directions may be selected, making the later interpretation of the incoming sound field much simpler than if the whole sound field (from many sources simultaneously) must be interpreted at once.

10 According to one aspect of the present invention there is provided a method of processing sound, the method comprising the steps of:

detecting sounds at at least two spaced detecting locations;

15 analysing the detected sounds to identify the angular relation between respective sound sources and said detecting locations;

permitting selection of an angular relation associated with a particular sound source; and

20 processing the detected sounds in response to said selection to highlight a stream of sound associated with said particular sound source.

25 Preferably, the angular relation between the respective sound sources is determined at least in part by reference to time differences between the sounds from the respective sound sources as detected at the spaced detecting locations. Most preferably, the angular relation between the respective sound sources is determined with

reference to time differences determined with reference to at least one feature of the detected sounds, the feature being selected from: pitch phase; amplitude modulation phase; and onset. Ideally, time differences are determined 5 with reference to a plurality of features of the detected sounds.

Preferably also, the angular relation between the respective sound sources is determined at least in part by reference to intensity differences between the sounds from 10 the respective sound sources as detected at the spaced detecting locations.

Preferably also, the sounds are detected at locations corresponding to the ear of a user, and the angular relation between the respective sound sources may be determined by reference to interaural time differences 15 (ITDs) and interaural intensity differences (IIDs).

Preferably also, the method further comprises selectively filtering the detected sounds from said spaced locations into a plurality of channels and then comparing 20 features of sound of each channel from one location with features of sound from a corresponding channel associated with the other location.

According to another aspect of the present invention, there is provided a method of processing sounds emanating 25 from a plurality of sound sources, the method comprising the steps of:

detecting sounds at at least two spaced detecting locations;

analysing the detected sounds to determine the angular relation between the respective sound sources and said detecting locations by reference to at least one of intensity differences between the sounds from the 5 respective sound sources as detected at the spaced detecting locations and time differences between the sounds from the respective sound sources as detected at the spaced detecting locations; and

streaming the sounds associated with at least one 10 sound source on the basis of said determined angular relation.

According to a further aspect of the present invention there is provided apparatus for processing sound, the apparatus comprising:

15 means for detecting sounds at at least two spaced detecting locations;

means for analysing the detected sounds to identify the angular relation between respective sound sources and said detecting locations;

20 means for permitting selection of an identified angular relation associated with a particular sound source; and

means for processing the detected sounds in response to said selection to highlight a stream of sound associated 25 with said particular sound source.

According to a still further aspect of the present invention there is provided apparatus for processing sounds emanating from a plurality of sound sources, the apparatus

comprising:

means for detecting sounds at at least two spaced detecting locations;

means for analysing the detected sounds to determine the angular relation between the respective sound sources and said detecting locations by reference to at least one of intensity differences between the sounds from the respective sound sources as detected at the spaced detecting locations and time differences between the sounds from the respective sound sources as detected at the spaced detecting locations; and

means for streaming the sounds associated with at least one sound source on the basis of said determined angular relation.

These and other aspects of the present invention will become apparent from the following description when taken in combination with the accompanying drawings, in which:

Figure 1 is a graph of interaural time delay (ITD) as a function of angle from a source of sound;

Figure 2 is a schematic overview, in block diagram form, of apparatus for processing sound in accordance with a preferred embodiment of the present invention;

Figure 2a illustrates the outputs from the bandpass filters of Figure 2 in greater detail;

Figure 3 is a block diagram illustrating the determination of interaural intensity difference (IID) in the apparatus of Figure 2;

Figures 4, 5 and 6 are block diagrams illustrating the

determination of interaural time differences (ITDs) based on onset in the apparatus of Figure 2;

Figures 7 to 11 are block diagrams illustrating the determination of ITDs based on amplitude modulated (AM) signals in the apparatus of Figure 2;

Figure 12 is a block diagram illustrating the determination of simple ITDs in the apparatus of Figure 2;

Figure 13 is a block diagram illustrating the display of IIDs and ITDs in the apparatus of Figure 2; and

Figures 14 to 16 are block diagrams illustrating the processing of the user's interaction with the display of Figure 13.

Reference is first made to Figure 1 of the drawings, in which interaural time delay (ITD) is graphed as a function of the angle between a source of sound and straight ahead, for an inter-ear separation (signal path difference) of 150mm. The source distance is assumed to be large compared to the distance between the ears. As will be described, the apparatus of the preferred embodiment of the present invention determines ITD by a number of different routes, which information is then utilised to allow the apparatus to be used to stream sounds for a user.

Figure 2 is an overview of apparatus 10 for processing sound in accordance with a preferred embodiment of the present invention. The Figure shows the two input transducers 12, 14 (Microphone L and Microphone R), and two multiple channel bandpass filters 16, 18. The microphones 12, 14 may be placed in the ear, or on the back of the ear,

or at any suitable point, separated by an appropriate distance. The microphones 12, 14 are of an omnidirectional type: the directivity of the system is not achieved through microphone directional sensitivity. Further, the 5 microphones are matched, though this is not crucial.

Each bandpass filter 16, 18 separates the incoming electrical signal from the respective microphone 12, 14 into a number of bands, as illustrated in Figure 2a. These bands may overlap, and have a broad tuning: that is, they 10 have a characteristic roughly similar to the bands found in the sensitivity analysis of real animal cochleae. As an approximation, they have a bandwidth of about 10% of the centre frequency at 6dB.

The bandpass filters 16, 18 are matched to each other. 15 In addition, the filters 16, 18 have a fixed and known delay characteristic, and the delay characteristic is the same (or very close to the same) for the two filters 16, 18.

Both bandpass filters 16, 18 will have the same number 20 of outputs: the precise number is not material, but the performance of the system improves as the number of filters increases.

From the bandpass filters 16, 18, the features of the signals of the individual channels are processed to provide 25 information on interaural intensity differences (IIDs) and interaural time differences (ITDs). The resulting information is presented to the user in a format which allows the user to identify and select sources of sounds,

based on the direction of the sound reaching the user. The signals from the channel or channels primarily associated with the selected source are then processed to suit the user's particular requirements, thereby effectively 5 streaming the sound from the selected source, and minimising the effect of sounds or "noise" from other sources.

The operation of the apparatus 10 will be described initially with reference to Figure 2, followed by more 10 detailed descriptions of the manner in which individual features of the sound are detected, analysed and processed.

In the illustrated embodiment, the outputs from the filters 16, 18 are subject to four different forms of analysis, the necessary hardware being presented in the 15 Figures in the form of blocks. Each form of analysis is described below briefly, in turn.

The intensity of sound in each channel is computed, at 20, 22, and the determined intensities from the two microphones 12, 14 compared on a channel-by-channel basis, 20 at 24, to provide a measure of interaural intensity (IID) for each channel. Each IID indicates a particular angle between the microphones 12, 14 and the source of sound, and this information is stored, at 26, and also relayed to an interactive display 28. As will be described, this display 25 receives similar inputs from the results of the other forms of analysis, to provide more complete information for the user. The user may then interact with the display to select a particular "angle", or more particularly may

select to be presented with sound from the source corresponding to that angle. The output from the display 28 is then relayed to the respective stores 26, from which details of the channels "contributing" to or indicative of 5 the selected angle are extracted and relayed to a resynthesis substation 30. The channel-related information being received from the stores 26 by the substation 30 is used to select and process signals directly from the filters 16, 18, which signals are then selectively 10 processed to highlight the appropriate channel inputs and to present the signals to the user in an appropriate form, for example by selectively amplifying the selected channels.

In addition to IID, the sensitivity and accuracy of 15 the apparatus is improved by detecting and processing interaural time differences (ITDs) for each channel by three different methods, as described briefly below. Firstly, onset is detected and computed for each microphone, at 32 and 33, and the ITD computed and gated, 20 at 34, for each channel. The resulting onset ITDs are stored, at 36, and relayed to the interactive display 28 and the resynthesis substation 30 in a somewhat similar manner to the IID as described above.

Secondly, the amplitude modulation (AM) for each 25 channel is detected and grouped, at 38 and 39, and the resulting information used to compute and gate AM ITD, at 40. Again the resulting AM ITDs are stored, at 42, and relayed to the interactive display 28 and the resynthesis

substation 30 in a somewhat similar manner to the IID and onset ITD as described above.

Thirdly, the signal phases for each channel are detected and grouped, at 44 and 46, and the resulting 5 information used to compute signal phase ITD, at 48. Again the resulting signal phase ITDs are stored, at 50, and relayed to the interactive display 28 and the resynthesis substation 30 in a somewhat similar manner to the IID, onset ITD and AM ITD as described above.

10 These operations will now be described in greater detail, firstly by reference to Figure 3 of the drawings, which illustrates the computation of the estimate of the angle from which the dominant incoming sound originates using the inter-aural intensity difference between the left 15 and right inputs, one estimate being made per channel.

20 The IID is computed repeatedly (for example, every 25ms) for each channel, at 25. The IID computed is then turned into an estimate of the angle of incidence of the sound, at 24, using an estimate of the head-related transfer function. Note that this function is itself a complex function of the frequency of the sound. The angles thus estimated for each channel are grouped together, at 27, and a number of estimates of the incident angle of sounds made. These are then sent to the display subsystem 25 28.

Figure 4 of the drawings shows the onset detector 32 and the onset clustering detector 32a. There is one onset detector for each of the bands produced by the bandpass

filters of figure 2. Those from on the left side are processed separately from those from on the right side.

Each onset detector 32, 33 detects onsets (sudden increases in energy) in a single channel. The output of the onset detector is, in this implementation, a pulse. This pulse is produced very quickly after the increase in energy begins: in addition, the delay before the pulse is produced is independent of the size of the onset. Another way of saying this is to say that the latency of the pulse is low, and independent of the onset size. The output of the onset detector is written onset(x,i), where x is either L or R (for left or right side), and i identifies the bandpass channel.

The onset detector 32 outputs from a single side are fed to the onset cluster detector 32a. There is one onset cluster detector for each microphone (i.e. for each side). The onset cluster detector 32a groups together those onsets which have occurred within a short time (taking into account the differences in delay time across the filter bank).

The output of the onset cluster detector 32a is an n-channel signal (where n is the number of bandpass bands). Each signal is, at any time, either 1 (signifying that this channel is currently part of an onset cluster) or 0 (signifying that this channel is not part of an onset cluster).

Figure 5 shows the left and right onset cluster signals being composed, at 34a, to form a composite onset

cluster signal. This composition 34a may take the form of a set of n AND gates or a set of OR gates.

Figure 6 shows the onset signals from the left and right from each channel being compared in time, and a value for the time differential computed. There are n such values computed. For channels for which there was no onset signal, no output (null) will be produced: similarly for channels in which the lowest value for the time difference between the left and right onset signal is too large (that is, has a value which could not be produced from any signal direction), no output (null) will be produced.

The values produced will be gated, at 34b, by the composite onset signal cluster signal. This signal will select clusters of onsets (generally one at a time). There are n outputs produced: each is either a value for the time differential, or null.

One onset cluster is produced at a time, and the onset store 36 stores the channel sets of recent grouped onsets, indexed by grouped ITD, that is according to the determined direction of sound for the channel.

Figure 7 and figure 8 show how the amplitude modulation is detected. The outputs from the bandpass filters 16, 18 are rectified and smoothed, at 60, and the rectified smoothed output is supplied to an AM mapping network 62. This implementation is based on Smith L. S., A one-dimensional frequency map implemented using a network of integrate-and-fire neurons, in ICANN 98, Volume 2, p991-995, Springer verlag 1998. This network 62 (as shown in

figure 8) has a number of excitatory neurons ( $m$ ) 64, and one inhibitory neuron 66 (shaded). The input to all the excitatory neurons is the same: the input to the inhibitory neuron is the (delayed) output of the excitatory neurons.

5 The excitatory neurons are arranged so that they are each particularly sensitive to amplitude modulation at some small range of frequencies. The effect of the network is that, for amplitude modulated input, one of the excitatory neurons (the mapping neurons) fires in phase with the amplitude modulation. In addition, the inhibitory neuron pulses whenever there is a sufficient amount of amplitude modulation.

10

15 The AM selection stage takes the output from all the excitatory neurons. This is gated by the output from the inhibitory neuron (so that null output is produced in the absence of amplitude modulated input). It reduces the pulse output to a single amplitude modulated channel, by selecting only the active excitatory neuron output. Additionally, it codes the identity of the excitatory

20 neuron producing this output: this supplies information on the frequency of the amplitude modulation.

25 Figure 9 shows the production of the table 68 used in grouping amplitude modulated signals. In order to group the amplitude modulation signals (so that they can be used to compute grouped ITDs) we produce a table with an entry of 1 for each output for each AM frequency output from a bandpassed channel, and 0 otherwise. Thus each row of the table may contain at most one 1 entry. If the same AM

frequency is found in more than one channel, then there will be columns with more than one 1 entry. The illustration shows a situation with 15 bandpassed channels, and 12 AM distinguishable AM frequency bands. In the table shown, bandpassed bands 2, 3, and 9 have found AM in AM channel 3, bandpassed bands 6, 7, 8, and 11 have found AM in AM channel 7, and bandpassed bands 10, 12, and 14 have found AM in AM frequency channel 11.

One table is produced for each side (left, right): a composite table is produced by ANDing the left and right tables.

Figure 10 illustrates how the columns of the table are used to gate the AM signals at 40, selecting only those with the same AM frequency for comparison, and generation of interaural time differences (ITDs). In the figure, we show a system with 15 bandpassed channels. The output of column 7 of the table above has been used to gate these signals. Thus, only bands 6, 7, 8, and 11 have been selected.

These pulse signals (which will be at the same frequency - namely frequency band 7, and whose pulse times reflect the phase of the amplitude modulation, that is the pulses are in phase with the amplitude modulation) are then fed in pairs (left, right) to circuitry 70 which computes the time difference between these signals. The values across the different selected bands are then processed at 72 (for example, averaged, or the modal or median value selected) to produce the AM time differential signal for

this AM frequency band.

An AM time difference signal will be produced for each nonzero column in figure 9: that is for each AM frequency band detected in both Left and Right channels.

5       Figure 11 shows how recent time difference signals produced for each nonzero column are stored: that is, the set of channels associated with each signal is stored, indexed by the (grouped) ITD, as input from interactive display 28.

10      Figure 12 shows how simple (ungrouped) ITDs are computed from the output of each pair (left, right) of bandpassed channels. This is achieved by using a phase-locked pulse generator 74, 75 (which may, for example, generate a pulse on each positive-going zero-crossing), and then calculating the time difference between these pulses. 15      For low frequencies, these estimates tend to be unreliable, and for high frequencies, they can be ambiguous. However, there is a range of medium frequencies for which good estimates can be made.

20      One time difference estimate will be produced for each (medium-frequency) channel. These may be grouped together prior to further usage.

25      Recent time difference estimates are stored at 50: that is, the channels associated with each grouped time difference estimate are stored, indexed by the time difference (ITD) itself.

Figure 13 shows the time difference signals from the three sources (onsets (Figure 6), amplitude modulation

(Figure 11) and waveform-synchronous processing (Figure 12)) are displayed on the display 28, in the form of a direction display. That is, the time differences between the left and right channels is interpreted as an angle.

5 The angles computed from the IIDs (figure 3) are also displayed.

In this embodiment, the display takes the form of a semicircle, because the system cannot distinguish between sounds from in front and behind; darker areas correspond to estimated directions of sound source. The user interacts with the display, selecting a particular direction (e.g. by touching the display) from which they wish to be presented with sounds. The display returns the angle selected, and this is then processed. A low power flat touch panel display (such as those used in colour portable computers) may be utilised.

Figure 14 illustrates how the signals controlling the signal to be presented to the user are generated from the information recovered from the interactive display 28, and the stored information at the onset store 36 (Figure 6), the AM difference signal store 42 (Figure 11) and the stored waveform-based time difference signal store 50 (Figure 12).

The angle output from the interactive display 28 is computed, at 76, from the user's interaction with the display. This is converted into an estimate of the IID and ITD, at 78 and 80, which sound from that direction would lead to. The channel contributions from the low, medium

and high frequency channels are normalised, at 82, 84, to provide mixing signals.

Figure 15 shows how the angle computed from the user's interaction with the display is used to index into the store 26 of low-frequency channel contributions, to estimate which of the low frequency (LF) channels gave rise to IIDs which were likely to have been produced by signals from that direction. This will use the head-related transfer function which is different at different frequencies.

Figure 16 shows how the final signal for representation to the user is generated. The mixing signals ControlLMix and ControlRMix are generated as described in figure 14, and control left and right channel mixers 86, 88. The final mixing of the signals for the two ears, in mixers 90, 92 is controlled by signals ControlLRMix, and these will depend on the nature of the user's hearing loss.

As noted, previously, the present invention is intended to mimic, to a certain extent, the processing of sounds in the early auditory system of a human (or other mammal), as discussed below. The input to the system comes from two microphones 12, 14 (L, for left, and R, for right in the figures), which are placed a distance apart. The microphones may be, for example, placed at the end of the auditory canal, or elsewhere on the pinna. Placing the microphones at the end of the auditory canal allows the pinna transfer characteristic to alter the relative strengths of different frequencies. If final presentation

is binaural, this information is useful to the user, allowing them to place the sound in space better (J. Blauert Spatial Hearing MIT Press, revised edition, 1996). The microphones 12, 14 transduce the acoustic signal into 5 a electrical signal. These electrical signals are amplified (maintaining the same frequency/phase response in both channels), and fed into identical bandpass filter banks 16, 18. These filter banks 16, 18 perform a similar task to that of the cochlea. Each of these filter banks 10 16, 18 produces a large number of outputs, one for each channel. These channel outputs are used as input to modules which emulate the onset detectors, waveform-synchrony detectors and amplitude modulation detectors of 15 the neurobiological system. However, not all channels will use all three modules.

Earlier work (L.S. Smith. Onset-based sound segmentation. In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, editors, Advances in Neural Information Processing Systems 8, pages 729--735. MIT Press, 1996) has 20 used onsets in different channels for sound segmentation: however, the integrate-and-fire neuron models used there have a latency (time from the sudden increase occurring to the neuron firing) which is dependent on the volume of the sound, and the speed of increase, unlike those of the real 25 onset cells intensity (J.S. Rothman, E.D. Young, and P.D. Manis. Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: Implications of a computational model. Journal of Neurophysiology,

70(6):2562--2583, 1993; J.S. Rothman and E.D. Young. Enhancement of neural synchronization in computational models of ventral cochlear nucleus bushy cells. *Auditory Neuroscience*, 2:47--62, 1996). The synthetic onset detector must have a very short, but constant latency: this latency needs to be constant over a wide range both of intensities and of rates of increase. Since onsets may be used in location of both pitched and unpitched sounds, each onset detector may receive input from a range of bandpassed channels. Unlike the biological system we use one precise onset detector per channel, rather than rely on population coding.

Waveform-synchrony is primarily of use at low to medium frequencies, as discussed earlier. The synthetic waveform-synchrony detector will provide an output at a specific part of the phase of the signal (for example, at each positive-going zero-crossing). For precise measurement of the ITD, there needs to be as little jitter as possible.

Amplitude modulation is primarily useful at medium to high frequencies. Note that effective use of AM is predicated on the bandpass filter having a wide-band response such as the response of the real cochlea. Again, the detector must provide an output at a particular point in the envelope, for example at peaks, and again, jitter needs to be minimised.

The display shows the azimuthal direction of the different incoming sounds (though not whether the sound is

ahead of or behind the user) as computed from IIDs and head-related transfer functions, and from ITDs. This on its own may be used to draw attention to features of the auditory environment. However, it may be rendered more 5 useful to the hearing impaired by permitting them to interact with it to select the information to be presented to them by the hearing aid itself. How this is best achieved in a particular application will depend on factors which will vary from user to user, such as whether they are 10 willing to use their hands to interact with the system, or would prefer to interact only by turning their heads.

Two main modes of sound selection are likely to be utilised. In the embodiment likely to be preferred by most users, the user turns to face the (known) source of the 15 sounds in which they are interested. The sounds to be selected are then those with low ITD and IID. In another embodiment, as described above a map of the incoming sounds is produced and displayed, and the user selects the sounds to be presented.

20 The information to be presented to the user may be presented monaurally or binaurally. The following discussion, as illustrated in figures 14-16, refers to binaural presentation. The result of the user's interaction with the interactive display is an angle,  $\theta$  between  $-\pi/2$  and  $+\pi/2$  or  $0$ , if the user requests only those 25 sources directly ahead. This angle is used to compute the expected IID and ITD for signals from that direction. In the illustrated embodiment, the ITD estimate is used to

index into the stored ITD/channel list for onsets, amplitude modulation, and waveform synchronous subsystems. For each channel, these three values are used to compute an estimate of that channel's contribution produced by sources from that direction. For each ear, these estimates are normalised (to a length of 1), and this vector (ControlLMix for the left ear, and ControlRMix for the right ear) used to control a mixer (Figure 16). This produces two multichannel outputs, OutDataL and OutDataR. These are used for medium and high frequencies: for lower frequencies, the same approach is made using the IID. The resultant output, OutData, is a multichannel signal, suitable for visual display. For auditory presentation, the signals in the different channels are added together in a manner which reflects the user's hearing deficit.

Exactly how the selected sounds are presented to the user depends very much on the sensory faculties of the user. If there is sufficient residual hearing, then selective amplification may be most suitable: if the residual hearing is restricted to particular frequency bands, then resynthesis may be more appropriate. It may also be possible to mix these two techniques. Alternatively, presentation may use the visual modality.

The selected sound, produced as outlined in Figures 14 to 16 may have some channels selectively amplified to make up for the hearing deficit. The resulting sound may be presented (a) monaurally, if there is only sufficient residual hearing in one ear, or (b) binaurally if there is

sufficient residual hearing in both ears. In this case, we would present the data from OutDataL to the left ear, and from OutDataR to the right ear.

Additionally, we can still use the ControlLRMix signal to alter the gain on the signals from the two ears.

Where the residual hearing is restricted to a small part of the auditory spectrum, or, indeed, where the presentation takes place through an implant, it may be more appropriate to resynthesize the sound to take advantage of whatever hearing is available. Again, the signal we start from is the OutData signal.

Where there is little or no residual hearing, the information from the sound is presented in one particular direction visually, utilising a colour display to present information about how the power of the sound is distributed over the spectrum.

One possibility (which does not use the interactive display, but displays all the incoming sound) is to choose the colour to match the ITD, and to make the intensity reflect the strength of the signal.

Alternatively, one may use the interactive display to select the direction of the sources to be presented, and use the colour of the display to show the presence and pitch of amplitude modulation, keeping white for non-amplitude modulated areas of the spectrum, and again using the intensity to show the signal strength. This would use information present in Figures 8 to 10, but not used in the auditory presentation.

The interactive system operates under strict real-time constraints. In addition, the system is ideally light, and wearable, and runs with a low power consumption. Most current sophisticated hearing aids use digital signal processing (DSP) techniques. DSP circuits are generally organised as reconfigurable fast parallel multipliers and adders. This is highly appropriate for convolution computation, and is highly effective for digital filtering. Non-linear operations are also possible on such circuits.

5 However, although DSP technology is very fast; it is not inherently parallel, and we wish to process multiple channels simultaneously. In addition, there is a speed/power tradeoff.

10

An alternative technology is subthreshold analog VLSI (C. Mead. Analog VLSI and Neural Systems. Addison-Wesley, 1989). This technology works at extremely low power levels, and this allows highly parallel circuits which operate at low power to be utilised. In addition, the exponential characteristic of one of the basic components, the transconductance amplifier, mirrors the characteristics of the biological system rather better than either digital on/off switches, or more linear analogue devices.

15

20 Sound detection may be through microphones. Alternatively, direct silicon transducers for pressure waves may be used. In the preferred embodiment there are two microphones, mounted on the user's ears (either behind the ear, or in the auditory canal). The microphones are omnidirectional: we need to receive signals from all

25

directions so we can estimate the directions of sound sources.

We describe below one possible neuromorphic implementation of some of the processing described above, though it is understood that this implementation is given by way of example only and is not intended to limit the scope of the invention. This processing takes place in stages. For the input from each ear, the first stage (after transduction) is cochlear filtering, and this is followed (in each bandpassed channel) by (in parallel) intensity computation, pitch phase detection, and envelope processing (that is amplitude modulation phase detection and onset detection). The results of this processing (for all channels, and for both ears) are used to generate ITD estimates for each feature type for each channel. This information is then used in determining what should be presented to the user.

The use of neuromorphic technology for real-time cochlear filtering was initially proposed by Lyon et al (R.F. Lyon and C. Mead. An analog electronic cochlea. IEEE Transactions on Acoustics, Speech and Signal Processing, 36(7):1119--1134, 1988) and has been extended by Lazzaro (J. Lazzaro and C. Mead Silicon modeling of pitch perception Proceedings of the National Academy of Sciences of the United States, 86(23):9597--9601, 1989), Liu and Andreou (W. Liu, A.G. Andreou, and Jr. M.H. Goldstein. Analog cochlear model for multiresolution speech analysis. In Advances in Neural Information Processing Systems 5,

pages 666--673, 1993, and W. Liu, A.G. Andreou, and Jr. M.H. Goldstein. Voiced speech representation by an analog silicon model of the auditory periphery. *IEEE Trans. Neural Networks*, 3(3):477--487, 1993), Watts (L. Watts *Cochlear Mechanics: Analysis and Analog VLSI* PhD thesis, California Institute of Technology, 1993), and more recently by Frangiere and van Schaik ( E. Fragniere, A. van Schaik, and E.A. Vittoz *Design of an analogue VLSI model of an active cochlea* *Analog Integrated Circuits and Signal Processing*, 12:19--35, 1997). The advantages of the neuromorphic solution are that it is inherently real-time, and low power, unlike DSP implementations. At present it is not yet possible to achieve as high a quality factor (Q) or as many stages as achieved by the human cochlea, but the most recent techniques (A. van Schaik. *Analogue VLSI Building Blocks for an Electronic Auditory Pathway*. PhD thesis, Ecole Polytechnique Federale de Lausanne, 1997) can provide 104 stages using a second order low-pass filter cascade.

Pitch phase detection in animals relies on population coding by spiking neurons which are more likely to spike at a particular phase of the movement of the basilar membrane. Neuromorphic implementations of this are discussed by Liu et al (W. Liu, A.G. Andreou, and Jr. M.H. Goldstein *Voiced speech representation by an analog silicon model of the auditory periphery* *IEEE Trans. Neural Networks*, 3(3):477--487, 1993) and in techniques by Van Schaik (A. van Schaik. *Analogue VLSI Building Blocks for an Electronic Auditory Pathway*. PhD thesis, Ecole Polytechnique Federale de

Lausanne, 1997), where a version of Meddis's hair cell model (M.J. Hewitt and R. Meddis An evaluation of eight computer models of mammalian inner hair-cell function, Journal of the Acoustical Society of America, 90(2):904--917, 1991) is implemented. In both these cases, both the tendency to synchronize with the input signal below about 4KHz, and the rapid and short-term adaptation are modelled.

5 However, if the aim is simply to encode the phase of the signal emanating from each bandpass filter, then a further available technique would be rectification followed by peak detection, or alternatively, simple positive-going zero crossing detection. Either of these can be easily accomplished using neuromorphic techniques. Lazzaro et al (J. Lazzaro and C.A. Mead. A silicon model of auditory 10 localization. Neural Computation, 1(1):47--57, 1989) have implemented neuromorphically a model of the barn owl's auditory localisation system using a detector sensitive to 15 zero-crossings of the derivative of the half-wave rectified bandpass filter output.

20 Although it would be possible for a neuromorphic system to retain waveform-synchronous operation at high frequencies, source direction detection is difficult because of the short period of these signals. Matching the peaks leads to ambiguity in the source direction. However, 25 if the result of bandpassing the signal at high frequencies is that there is amplitude modulation at a lower frequency, then the difference in the phase of the modulation between the two detectors may be used. Neuromorphic detection of

amplitude modulation (modelling stellate cells in the cochlear nucleus) is discussed by Van Schaik (A. van Schaik Analogue VLSI Building Blocks for an Electronic Auditory Pathway PhD thesis, Ecole Polytechnique Federale de 5 Lausanne, 1997) in the context of periodicity extraction.

Although the same techniques could be used for ITD estimation, it is perhaps simpler to low-pass filter the waveform-synchronous phase detector, and generate a pulse on each peak (or on each positive-going zero-crossing).

10 Neuromorphic implementation of the onset detector may be achieved using a neuromorphic spiking neuron.

15 Since there are three independent techniques for ITD computation in each channel (although amplitude modulation would not be used below about 1KHz, and waveform synchrony would not be used above about 4KHz), we are liable to have both a number of estimates at different parts of the spectrum, and even a number of estimates at each part of the spectrum. There may be many sound sources at any one time, so that all these estimates may well be correct.

20 A mixture of subthreshold analogue, supra-threshold analogue and digital techniques may be applied to the production of the neuromorphic implementation of the control signal generation and of the mixers.

25 What is to be presented will be produced from the OutData signal (or OutDataL and OutDataR signals in the case of binaural presentation). Auditory presentation technology may, for example, utilise remote generation of the signal, and transmission of the signal to the in-ear

transducers by wireless technology. In addition, it may be necessary to adjust the spectral energy distribution and compress the signal to take best advantage of the residual hearing present.

5 Noting that the bandpass characteristic of current neuromorphic filters is not as sharp as is preferred we may counterbalance this by (i) selectively amplifying those channels for which the chosen ITD is most strongly represented and (ii) subtracting the content of those  
10 channels in which the ITD chosen is under-represented.

It will be understood that the embodiments of the invention herein before described are given by way of example only and are not meant to limit the scope thereof in any way.

CLAIMS

1. A method of processing sound, the method comprising the steps of:

detecting sounds at at least two spaced detecting locations;

analysing the detected sounds to identify the angular relation between respective sound sources and said detecting locations;

permitting selection of an angular relation associated with a particular sound source; and

processing the detected sounds in response to said selection to highlight a stream of sound associated with said particular sound source.

2. The method of claim 1, wherein the angular relation between the respective sound sources is determined at least in part by reference to time differences between the sounds from the respective sound sources as detected at the spaced detecting locations.

3. The method of claim 2, wherein the angular relation between the respective sound sources is determined with reference to time differences determined with reference to at least one feature of the detected sounds, the feature being selected from: waveform phase; amplitude modulation phase; and onset.

4. The method of claim 3, wherein time differences are determined with reference to a plurality of features of the detected sounds.

5 5. The method of any of the preceding claims, wherein the angular relation between the respective sound sources is determined at least in part by reference to intensity differences between the sounds from the respective sound sources as detected at the spaced detecting locations.

10 6. The method of any of the preceding claims, wherein the sounds are detected at locations corresponding to the ears of a user.

7. The method of claim 6, wherein the angular relation between the respective sound sources is determined at least in part by reference to interaural time differences (ITDs).

15 8. The method of claims 6 or 7, wherein the angular relation between the respective sound sources is determined at least in part by reference to interaural intensity differences (IIDs).

20 9. The method of any of the preceding claims, further comprising selectively filtering the detected sounds from said spaced locations into a plurality of channels and then comparing features of sound of each channel from one location with features of sound from a corresponding

channel associated with the other location.

10. The method of claim 9 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to waveform phase of the detected sounds and wherein the time differences are grouped by clustering values of the time differences.

10 11. The method of claim 9 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to onset of the detected sounds and wherein onsets are grouped monaurally prior to determination of the time differences.

15 12. The method of claim 9 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to amplitude modulation phase of the detected sounds and wherein the amplitude modulation channels are grouped by amplitude modulation frequency prior to determination of the time differences.

20 13. A method of processing sounds emanating from a plurality of sound sources, the method comprising the steps of:

detecting sounds at at least two spaced detecting locations;

analysing the detected sounds to determine the angular relation between the respective sound sources and said detecting locations by reference to at least one of intensity differences between the sounds from the 5 respective sound sources as detected at the spaced detecting locations and time differences between the sounds from the respective sound sources as detected at the spaced detecting locations; and

streaming the sounds associated with at least one 10 sound source on the basis of said determined angular relation.

14. Apparatus for processing sound, the apparatus comprising:

means for detecting sounds at at least two spaced 15 detecting locations;

means for analysing the detected sounds to identify the angular relation between respective sound sources and said detecting locations;

means for permitting selection of an identified 20 angular relation associated with a particular sound source; and

means for processing the detected sounds in response to said selection to highlight a stream of sound associated with said particular sound source.

25 15. The apparatus of claim 14, wherein said analysing means includes means for determining the angular relation

between the respective sound sources by determining time differences between the sounds from the respective sound sources as detected at the spaced detecting locations.

16. The apparatus of claim 15, wherein said analysing means includes means for determining the angular relation between the respective sound sources by determining time differences between the sounds from the respective sound sources as detected at the spaced detecting locations by reference to at least one of waveform phase; amplitude modulation phase and onset.

17. The apparatus of claim 16, wherein said analysing means includes means for determining the angular relation between the respective sound sources by determining time differences between the sounds from the respective sound sources as detected at the spaced detecting locations by reference to a plurality of features of the detected sounds.

18. The apparatus of any of claims 13 to 17, wherein said analysing means includes means for determining the angular relation between the respective sound sources by determining intensity differences between the sounds from the respective sound sources as detected at the spaced detecting locations.

19. The apparatus of any of claims 13 to 18, wherein the

apparatus is a hearing aid and said means for detecting sounds are adapted for positioning at locations corresponding to the ears of a user.

20. The hearing aid of claim 19, wherein said analysing means includes means for determining the angular relation between the respective sound sources by determining interaural time differences (ITDs) between the sounds from the respective sound sources.

10 21. The hearing aid of claims 19 or 20, wherein said analysing means includes means for determining the angular relation between the respective sound sources by determining interaural intensity differences (IIDs) between the sounds from the respective sound sources.

15 22. The apparatus of any of claims 14 to 21, further comprising means for selectively filtering the detected sounds from said spaced locations into a plurality of channels, the analysing means comprising means for comparing features of sound of each channel from one location with features of sound from a corresponding 20 channel associated with the other location.

23. The method of claim 22 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to waveform phase of the detected sounds and wherein the time

differences are grouped by clustering values of the time differences.

24. The method of claim 22 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to onset of the detected sounds and wherein onsets are grouped monaurally prior to determination of the time differences.

25. The method of claim 22 when the angular relation between the respective sound sources is determined with reference to time differences determined with reference to amplitude modulation phase of the detected sounds and wherein the amplitude modulation channels are grouped by amplitude modulation frequency prior to determination of the time differences.

15 26. Apparatus for processing sounds emanating from a plurality of sound sources, the apparatus comprising:

means for detecting sounds at at least two spaced detecting locations;

20 means for analysing the detected sounds to determine the angular relation between the respective sound sources and said detecting locations by reference to at least one of intensity differences between the sounds from the respective sound sources as detected at the spaced detecting locations and time differences between the sounds from the respective sound sources as detected at the spaced

25

detecting locations; and

means for streaming the sounds associated with at least one sound source on the basis of said determined angular relation.

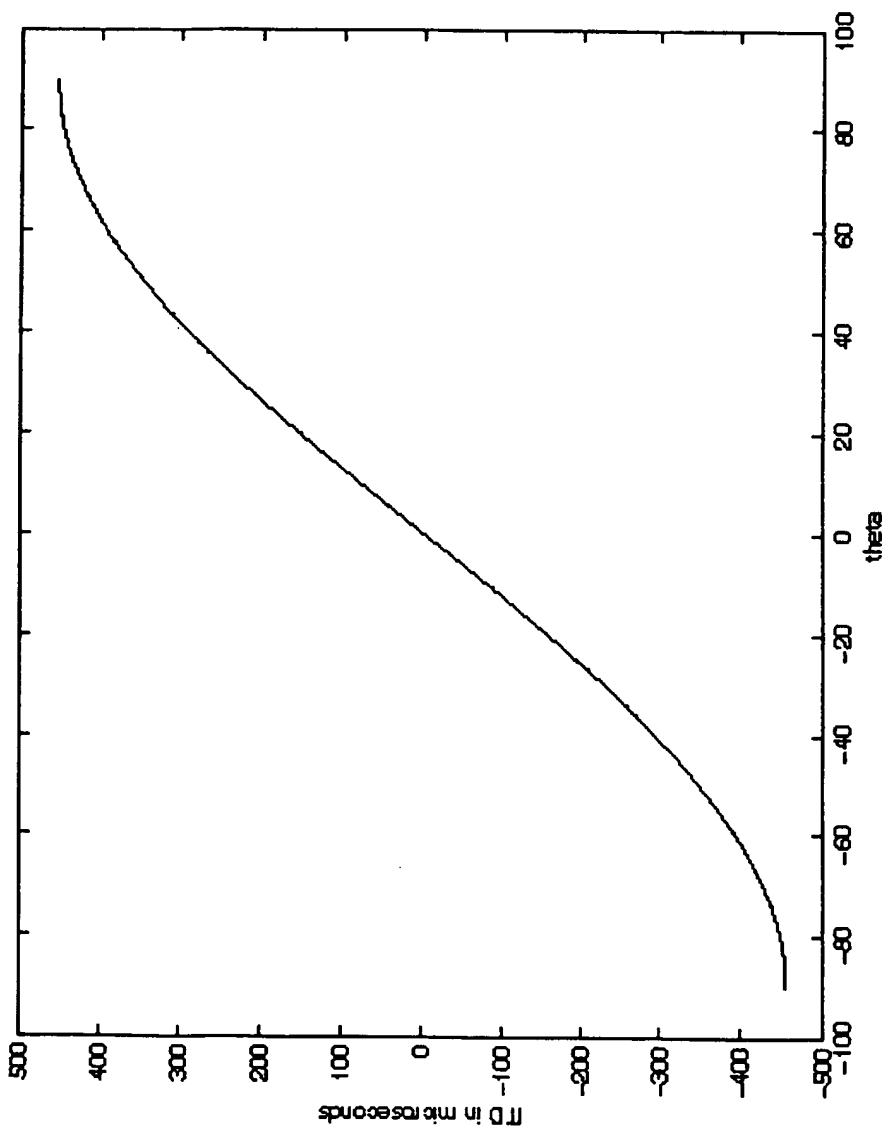


Figure 1.

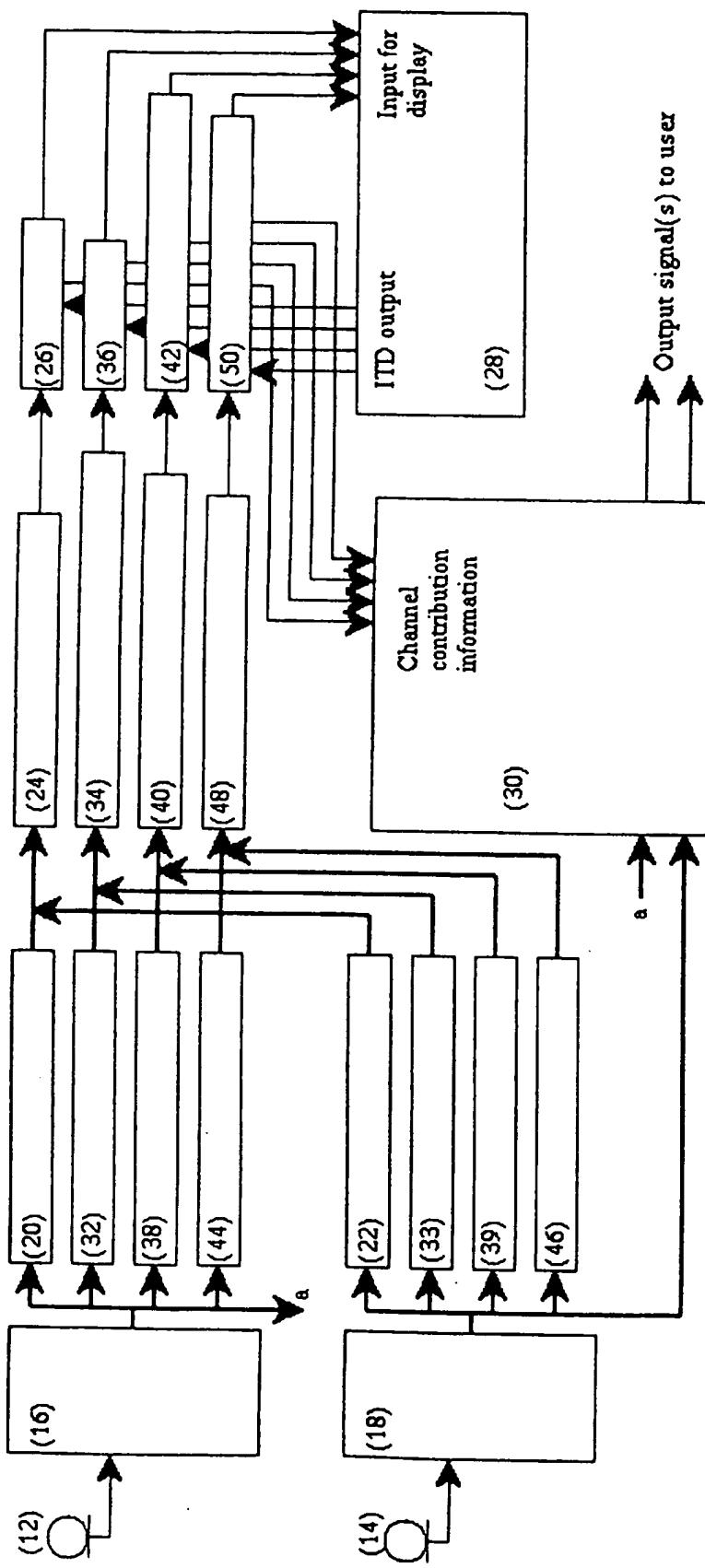
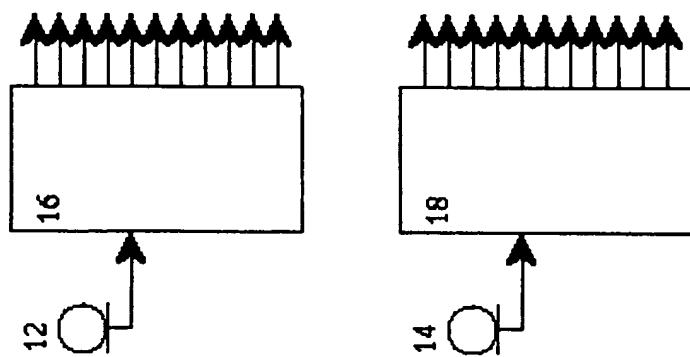


Figure 2.

Figure  
2A

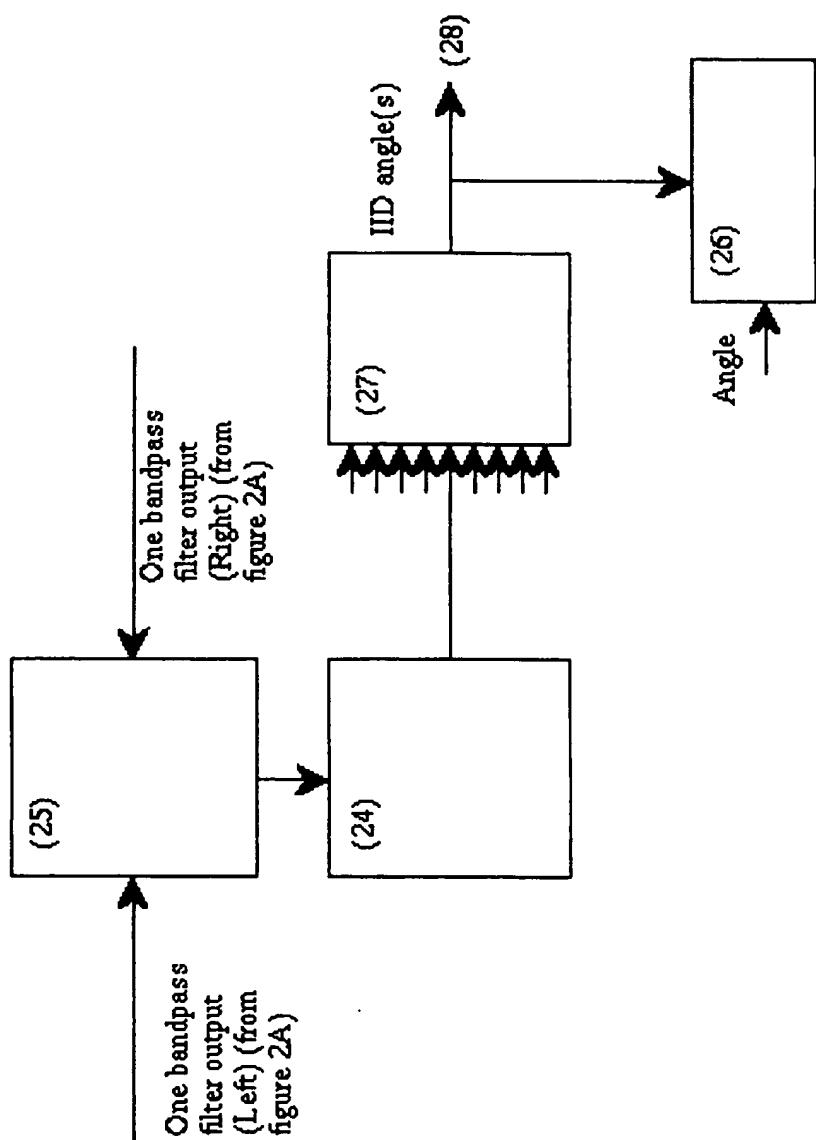


Figure 3

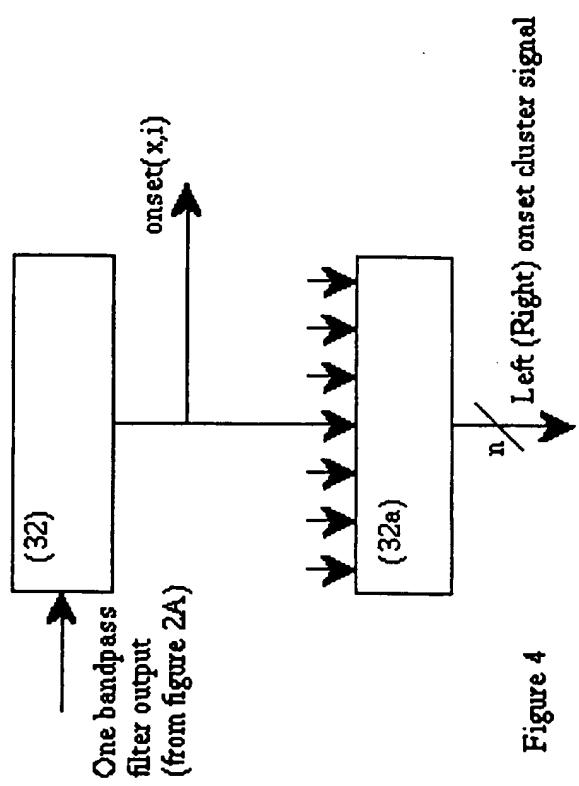


Figure 4

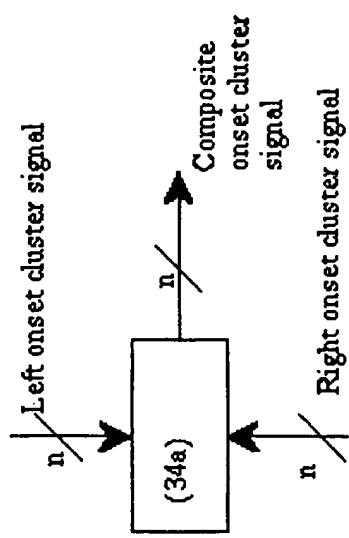


Figure 5

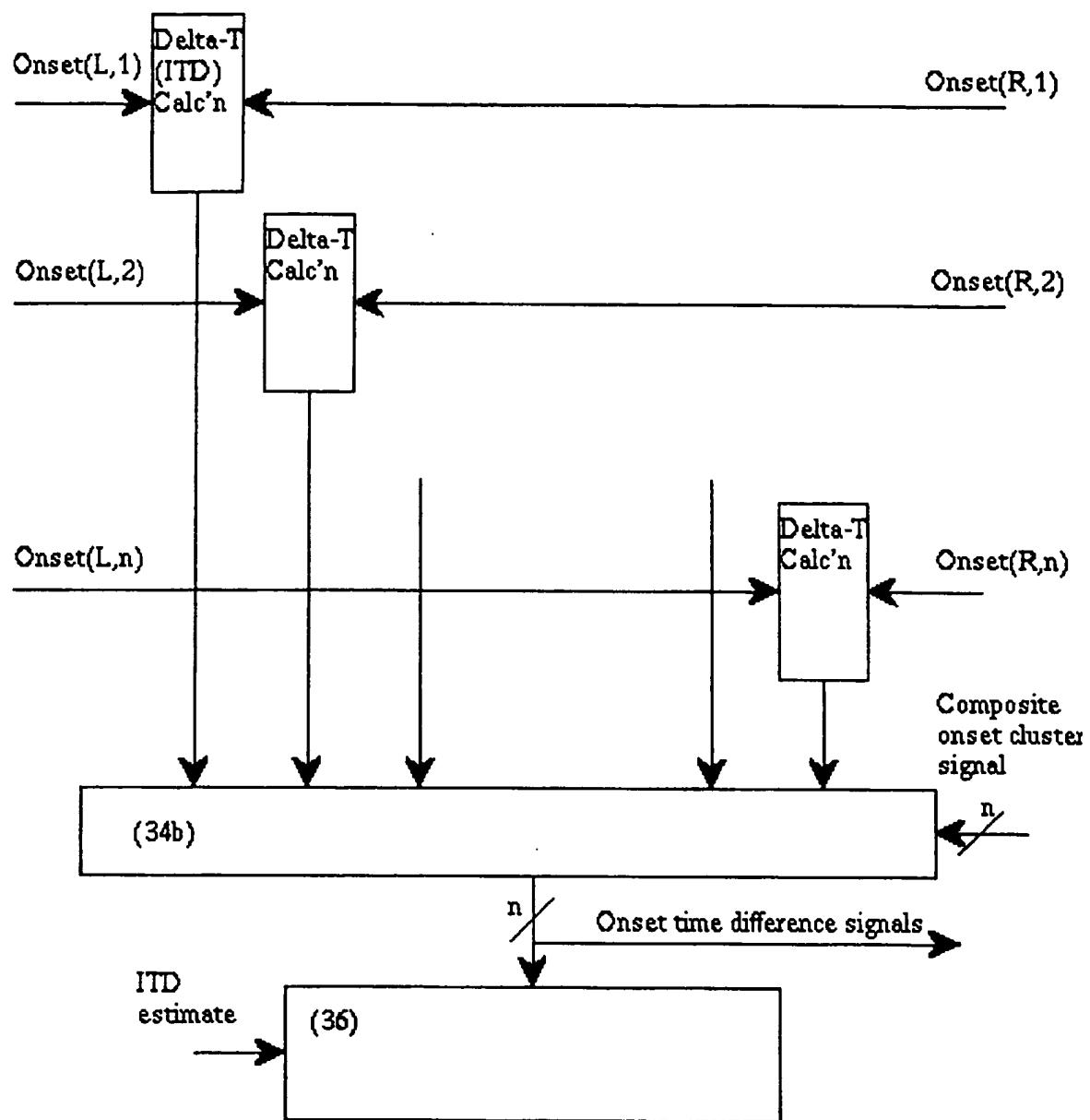


Figure 6

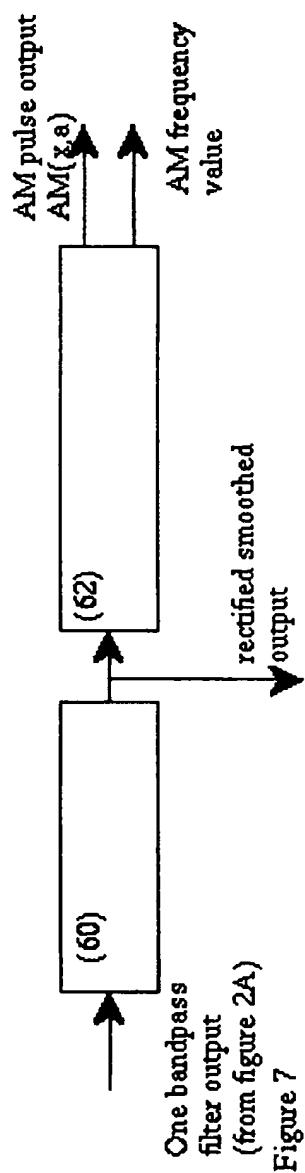


Figure 7

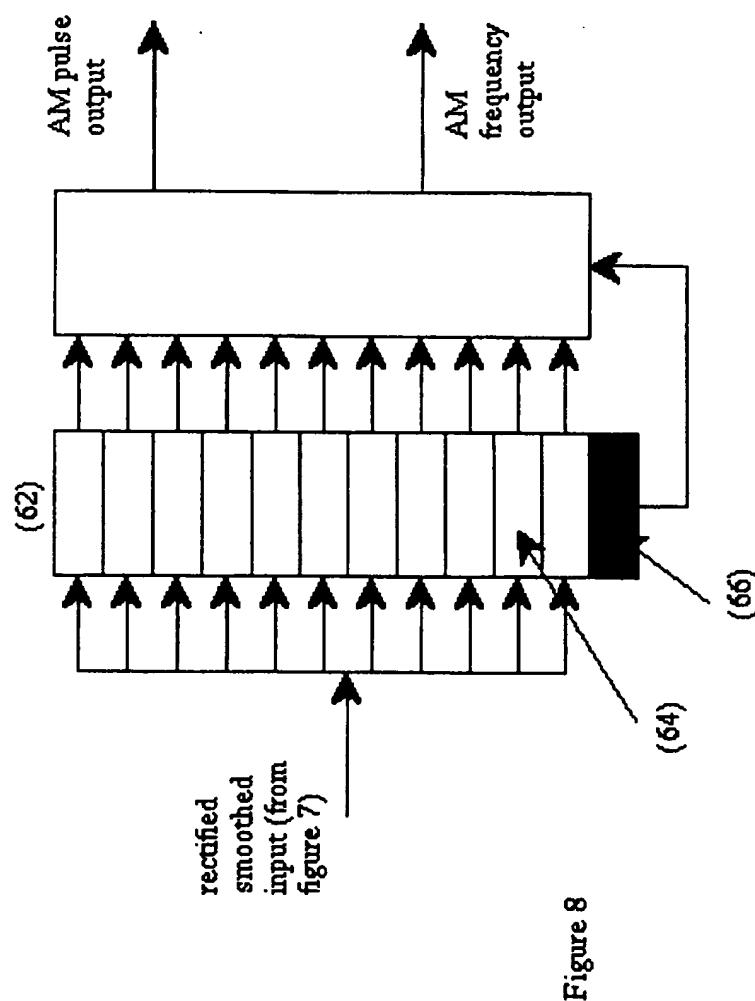


Figure 8

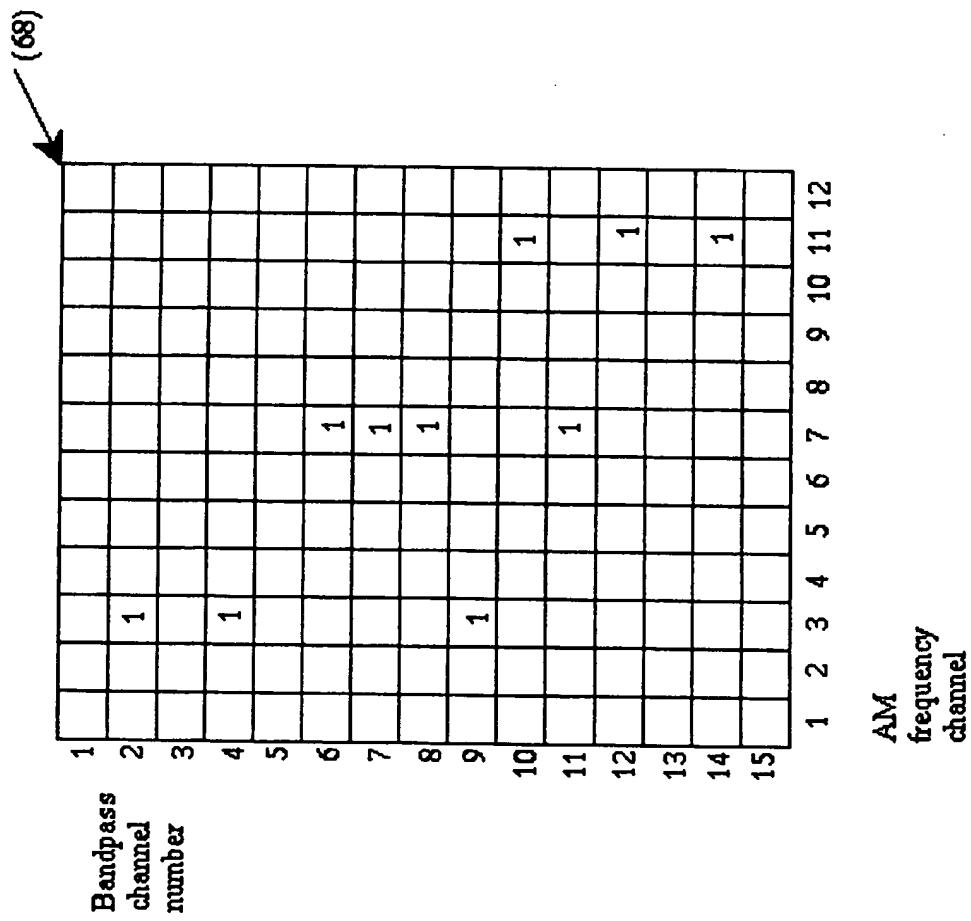


Figure 9

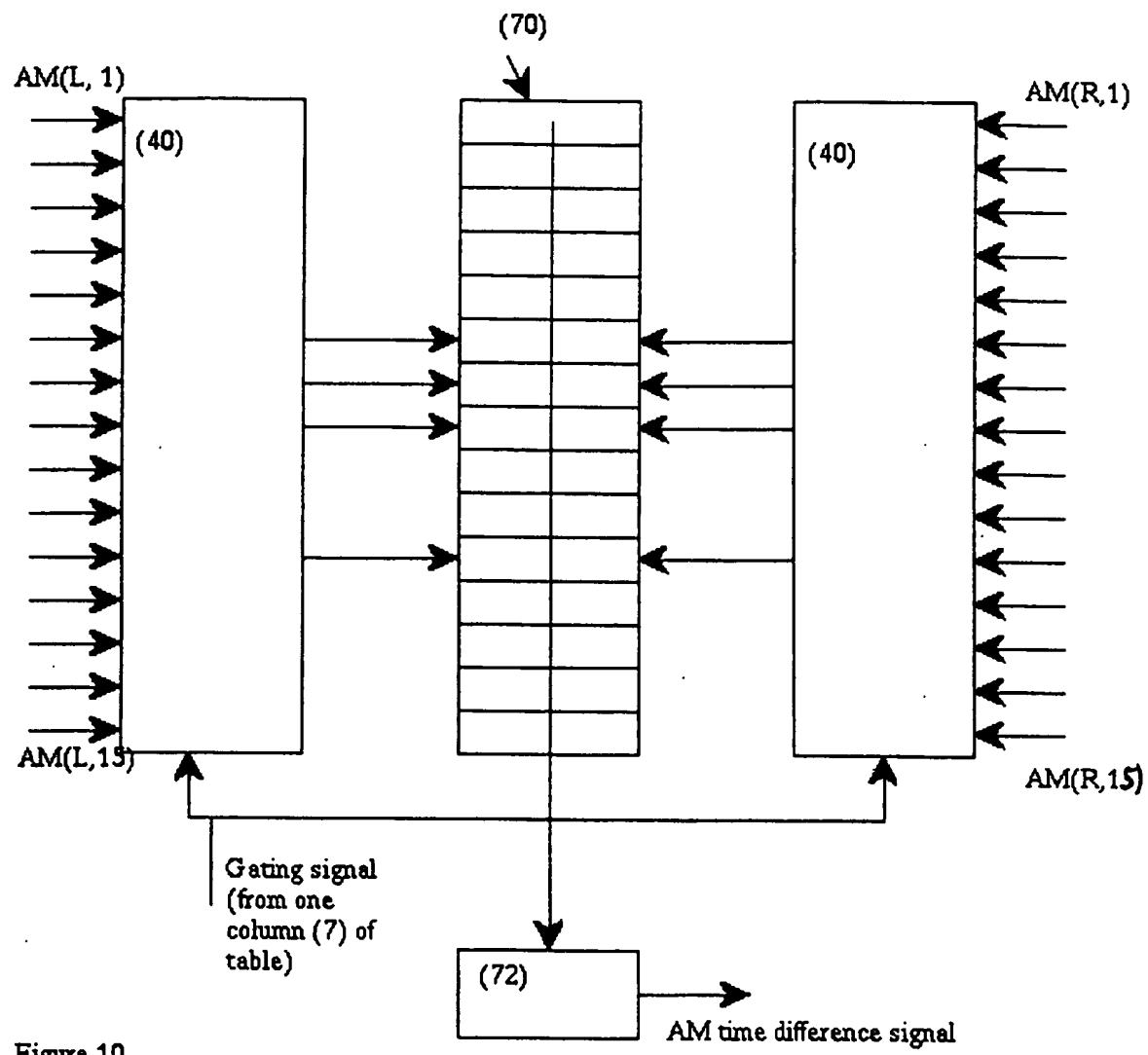


Figure 10

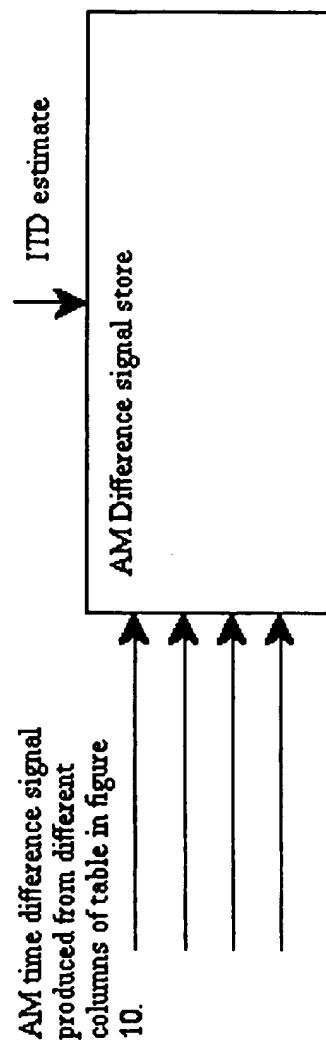


Figure 11

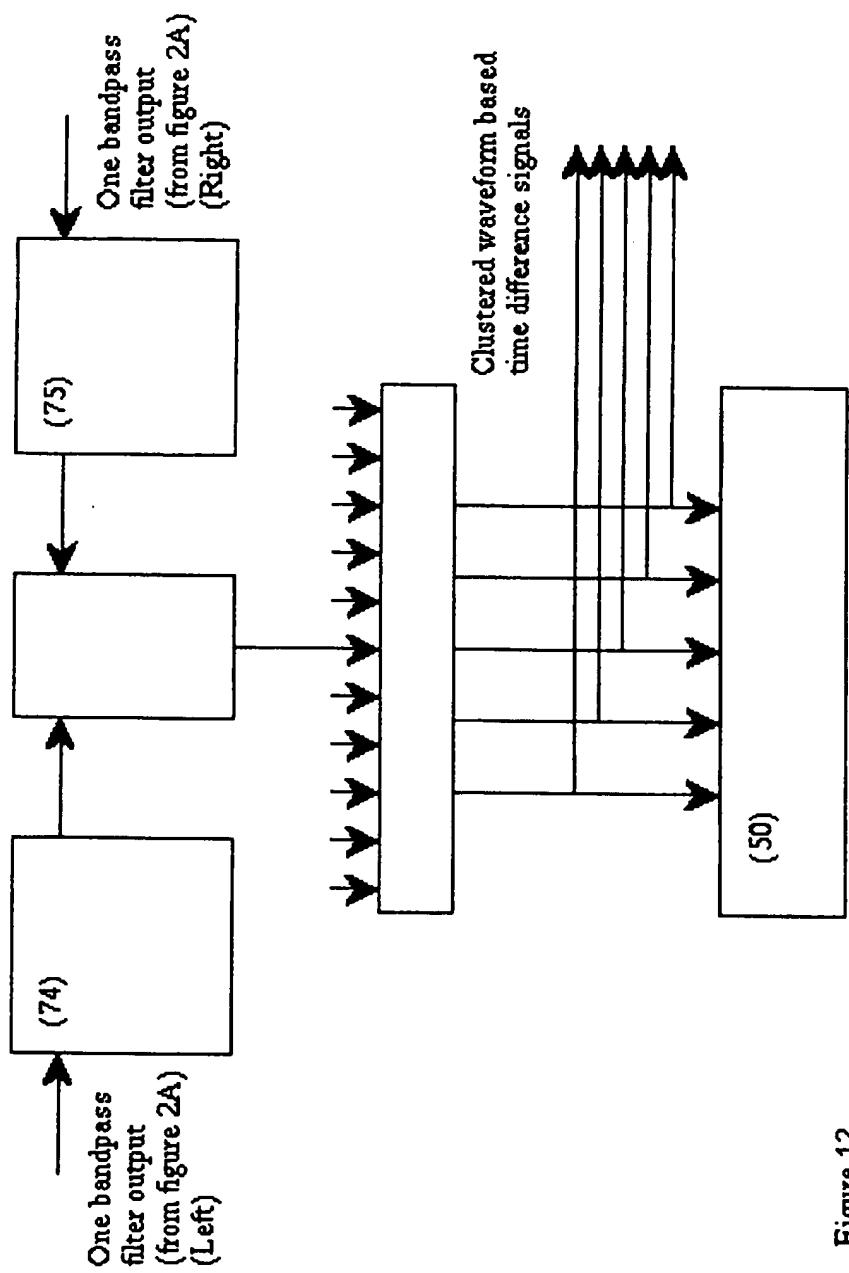


Figure 12

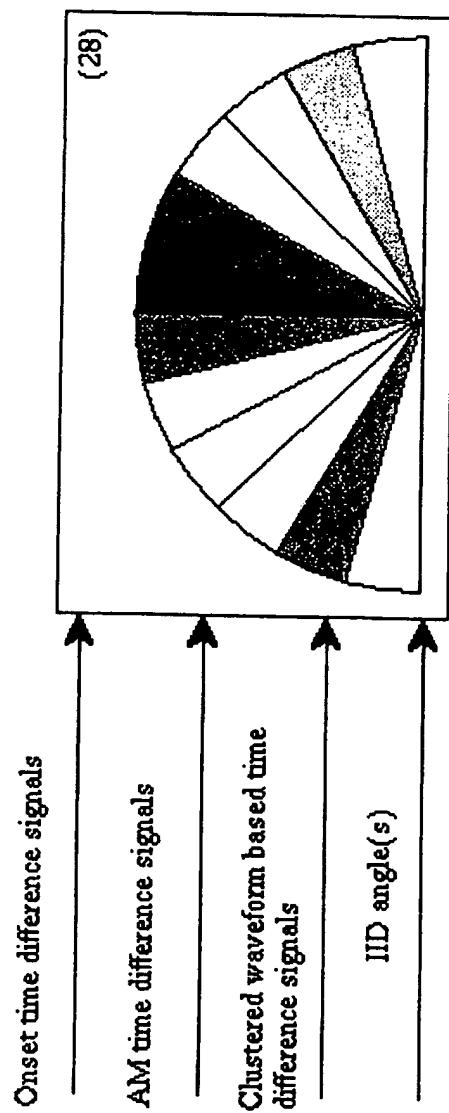


Figure 13

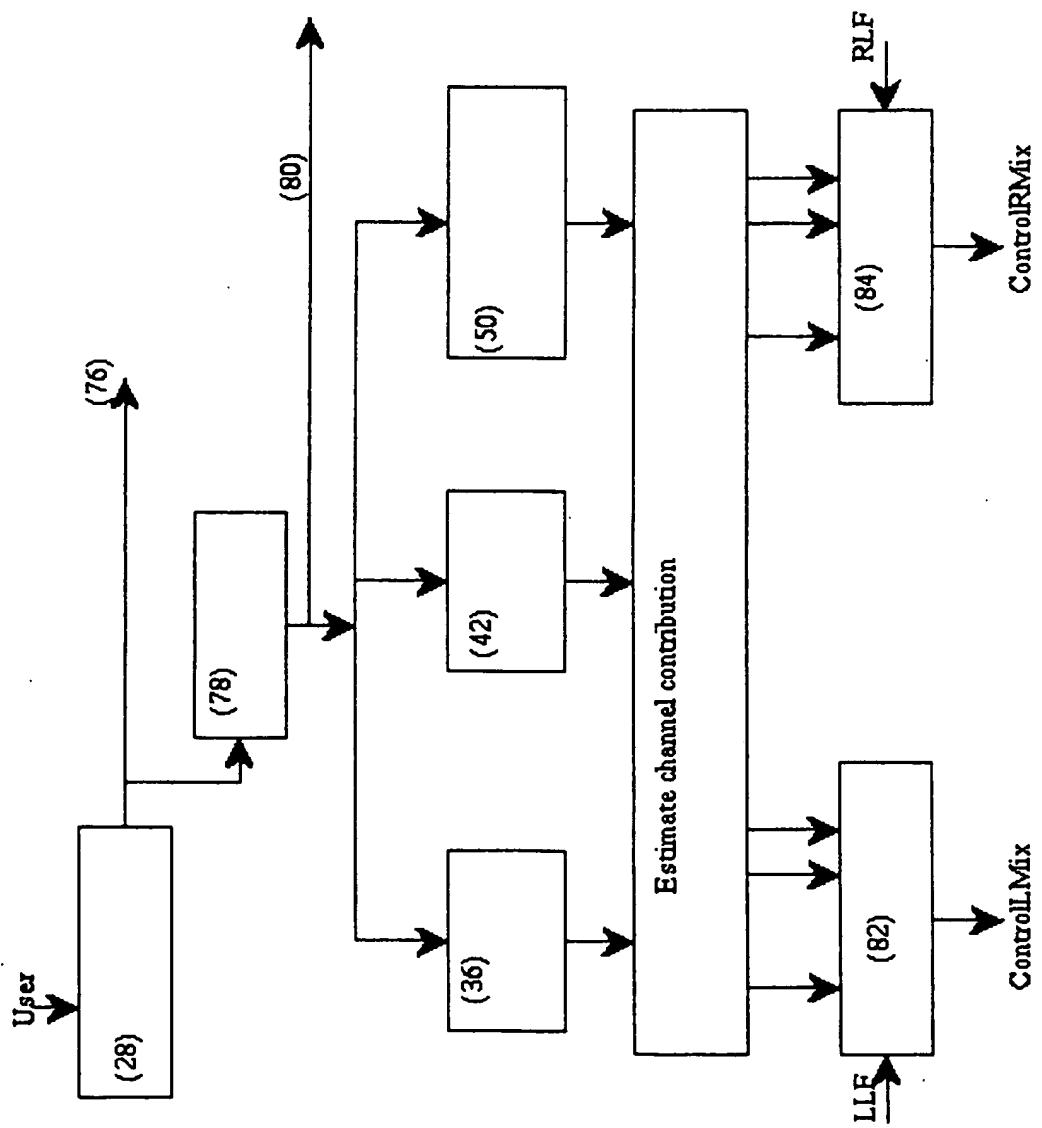


Figure 14

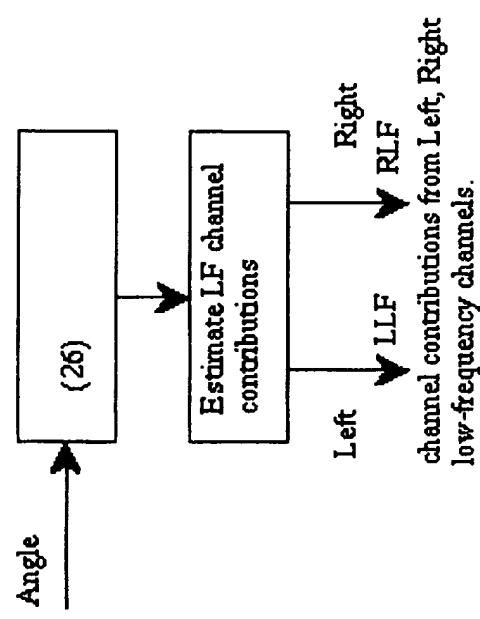


Figure 15

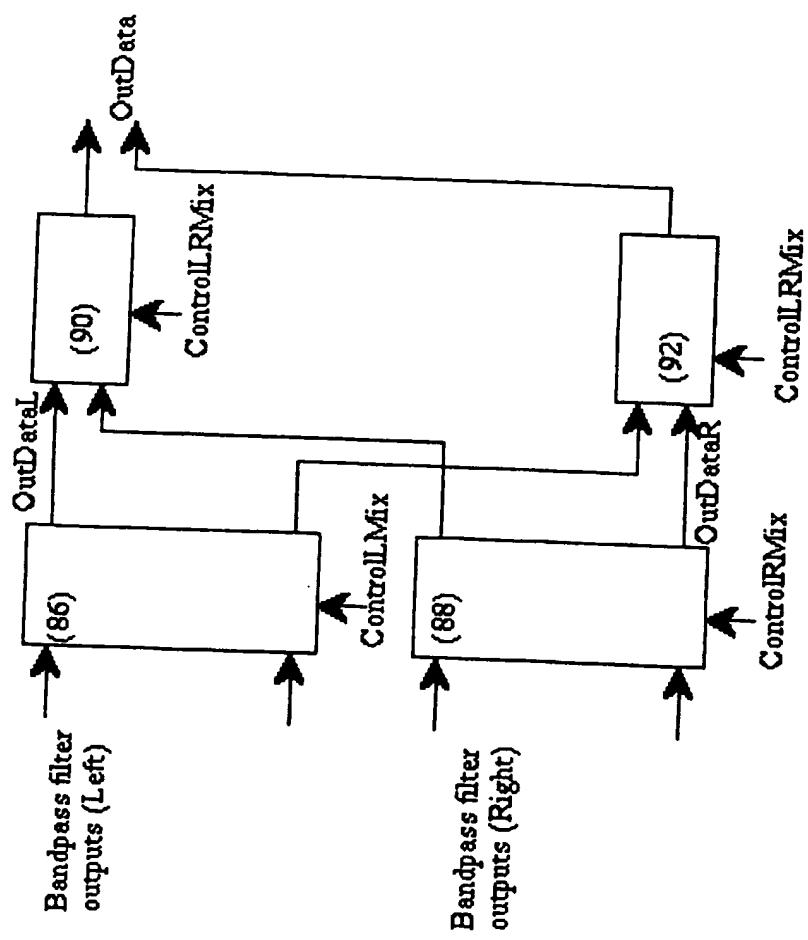


Figure 16

# INTERNATIONAL SEARCH REPORT

International Application No

P.. / GB 99/02063

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 H04R25/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 H04R G01S

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PATENT ABSTRACTS OF JAPAN vol. 1995, no. 5, 30 June 1995 (1995-06-30) & JP 07 039000 A (KAZUMOTO SUZUKI), 7 February 1995 (1995-02-07) abstract	1-5, 9, 10, 13-18, 22, 23, 26
Y		6-8, 19-21
A		11, 12, 24, 25
Y	---	6-8, 19-21
	US 4 904 078 A (GORIKE) 27 February 1990 (1990-02-27) column 2, line 35-38 column 4, line 42 -column 5, line 5 column 6, line 3-37 ---	---
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 November 1999	18/11/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl Fax: (+31-70) 340-3016	Authorized officer  Zanti, P

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/02063

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 757 932 A (LINDEMANN ET AL.) 26 May 1998 (1998-05-26) cited in the application column 1, line 34-65 column 2, line 14-19 column 3, line 5 -column 5, line 18 column 6, line 1 -column 7, line 17 column 8, line 6-62 ---	6-8, 19-21
Y	PATENT ABSTRACTS OF JAPAN vol. 1998, no. 1, 30 January 1998 (1998-01-30) & JP 09 247800 A (MATSUSHITA), 19 September 1997 (1997-09-19) abstract ---	6-8
Y	DE 15 66 857 A (F.KRUPP GMBH) 30 April 1970 (1970-04-30) page 1, line 5-7 page 2, line 26 -page 3, line 15 page 3, line 22 -page 5, line 17 ---	6-8
X	PATENT ABSTRACTS OF JAPAN vol. 1997, no. 3, 31 March 1997 (1997-03-31) & JP 08 285674 A (HIRATA TAKAYOSHI), 1 November 1996 (1996-11-01) abstract -----	1-4, 14-17
A		13,26

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No

PCT/GB 99/02063

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
JP 07039000	A 07-02-1995	NONE		
US 4904078	A 27-02-1990	AT 383428 B		10-07-1987
		AT 97284 A		15-11-1986
		DE 3507397 A		26-09-1985
		JP 1581533 C		11-10-1990
		JP 2007050 B		15-02-1990
		JP 60213921 A		26-10-1985
US 5757932	A 26-05-1998	US 5479522 A		26-12-1995
		AU 7248796 A		30-04-1997
		EP 0855130 A		29-07-1998
		NO 981559 A		08-06-1998
		WO 9714268 A		17-04-1997
JP 09247800	A 19-09-1997	NONE		
DE 1566857	A 30-04-1970	NONE		
JP 08285674	A 01-11-1996	NONE		

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**